

Automatic Derivation of Morphological, Syntactic, and Semantic Meaning from a Natural Language System Using a Monte Carlo Markov Chain Process

Abstract

A method for deriving the morphology, syntax, and semantics of a language system (comprised of untagged free text) is presented. The concept of the "language object" (a unique data structure containing information concerning the behavior of a given segment of the input language) is introduced, and is shown to be useful in the analysis of a language system when utilized by a Monte Carlo Markov Chain rule engine to discern probabilities of various language rules and the existence of various "language objects." This process of positing and testing language objects and rules functions on morphologic, syntactic and semantic levels, building a comprehensive understanding of language use and structure from base elements up to the complex systems of human expression.